

中華民國經濟部智慧財產局

INTELLECTUAL PROPERTY OFFICE
MINISTRY OF ECONOMIC AFFAIRS
REPUBLIC OF CHINA

茲證明所附文件，係本局存檔中原申請案的副本，正確無訛，
其申請資料如下：

This is to certify that annexed is a true copy from the records of this
office of the application as originally filed which is identified hereunder:

申請日：西元 2003 年 09 月 12 日
Application Date

申請案號：092125187
Application No.

申請人：財團法人工業技術研究院
Applicant(s)

局長
Director General

蔡練生

發文日期：西元 2003 年 11 月 12 日
Issue Date

發文字號：09221141990
Serial No.

發明專利說明書

(本說明書格式、順序及粗體字，請勿任意更動，※記號部分請勿填寫)

※ 申請案號：

※ 申請日期：

※IPC 分類：

壹、發明名稱：(中文/英文)

連續語音自動切音及驗證之方法及系統

An automatic speech segmentation and verification system and its method

貳、申請人：(共 1 人)

姓名或名稱：(中文/英文)

財團法人工業技術研究院

Industrial Technology Research Institute

代表人：(中文/英文) 翁政義 / Cheng-I Weng

住居所或營業所地址：(中文/英文)

新竹縣竹東鎮中興路四段一九五號

No. 195, Sec. 4, Chung Hsing Rd., Chutung, Hsinchu

國籍：(中文/英文) 中華民國 / R.O.C.

參、發明人：(共 3 人)

姓名：(中文/英文)

1. 郭志忠 / KUO CHIH-CHUNG

2. 郭啟祥 / KUO CHI-SHIANG

3. 陳昭宏 / CHEN JAU-HUNG

住居所地址：(中文/英文)

1. 新竹市東區新莊里 10 鄰新莊街 87 號 4 樓

4 Fl., No. 87, Hsin-Chuang St., Hsinchuang Li, East Area, Hsinchu

2. 台南縣永康市正強里 25 鄰中正路 127 巷 22 號

No. 22, Lane 127, Chung-Cheng Rd., Cheng Ching Li, Yung Kang City, Tainan Hsien

3. 新竹市東區金山里 20 鄰光復路 1 段 89 巷 123-2 號 3 樓之 3

3Fl.-3, No.123-2, Lane 89, Sec. 1, Kuang-Fu Rd., Kin-Shan Li, East Area, Hsinchu

國籍：(中文/英文) 1.2.3. 中華民國 / R.O.C.

肆、聲明事項：

☒ 本案係符合專利法第二十條第一項 ☒ 第一款但書或 ☐ 第二款但書規定之期間，其日期為：2003 年 9 月 1-4 日。

◎ 本案申請前已向下列國家（地區）申請專利 ☐ 主張國際優先權：

【格式請依：受理國家（地區）；申請日；申請案號數 順序註記】

1. 無

2.

3.

4.

5.

☐ 主張國內優先權（專利法第二十五條之一）：

【格式請依：申請日；申請案號數 順序註記】

1.

2.

☐ 主張專利法第二十六條微生物：

☐ 國內微生物 【格式請依：寄存機構；日期；號碼 順序註記】

☐ 國外微生物 【格式請依：寄存國名；機構；日期；號碼 順序註記】

☐ 熟習該項技術者易於獲得，不須寄存。

伍、中文發明摘要：

本發明係有關於一種連續語音自動切音及驗證之方法及系統，用以對大語料庫所使用之合成語音單元進行自動切音驗證及語音驗證。首先接收一參照已知文字腳本所錄製而成的錄音語料檔；接著由單元切割器根據已知文字腳本，對錄音語料檔中的語音段進行切割，以切割出待測語音單元；之後，將交由切音驗證器以根據各種音節型態的語音參數統計資料，來檢查待測語音單元的切音結果是否正確；並使用語音驗證器以透過語音驗證模組來檢查待測語音單元是否被正確錄製；最後，單元篩選器將結合切音及語音驗證方法的可靠度量測值，來決定接受或拒絕待測語音單元。

陸、英文發明摘要：

An automatic speech segmentation and verification system and its method is disclosed, which has a known text script and a recorded speech corpus corresponding to the known text script. A speech unit segmentor segments the recorded speech corpus into N test speech unit segments referring to the phonetic information of the known text script. Then, a segmental verifier is applied to obtain a segmental confidence measure for verifying the correctness of the cutting points of test speech unit segments. A phonetic verifier obtains a phonetic confidence measure by using verification models for verifying whether the recorded speech corpus is correctly recorded. Finally, a speech unit inspector integrates the segmental confidence measure and the phonetic confidence measure to determine whether the test speech unit segment is accepted or not.

柒、指定代表圖：

(一)本案指定代表圖為：圖(1)。

(二)本代表圖之元件代表符號簡單說明：

電腦系統 1	資料庫 11
單元切割器 12	待測語音單元 121
切音驗證器 13	語音驗證器 14
單元篩選器 15	切音可靠度 CMS
語音可靠度 CMV	已知文字腳本 K
錄音語料檔 R	

捌、本案若有化學式時，請揭示最能顯示發明特徵的化學式：

無

玖、發明說明：

【發明所屬之技術領域】

本發明係有關語音合成之技術領域，尤指一種連續語音自動切音及驗證之方法及系統。

5

【先前技術】

按，在語音合成技術領域中，由於使用大語料庫 (speech corpus) 之電腦語音合成方法可取得具有高品質以及高流暢度的語音合成，其效果遠優於傳統僅採用訊號處理方式藉由調整基本音節的語音長度 (duration) 和音調高低 (pitch) 所合成的語音，因此使用大語料庫之語音合成已成為現今語音合成系統的主流。語音合成系統的成功與否取決於許多因素，諸如合成單元的數量多寡、錄音語料的品質優劣、選取合成單元的方法、自然韻律的產生、以及
10 合成單元的串接等。而隨著電腦處理速度的提升、與大容量記憶體和硬碟的普及，使得習知大語料庫的語音合成系統已可儲存數以萬計的合成單元，並能在合成的時候，即時從中選取出合適的合成單元。
15

於習知使用大語料庫的電腦語音合成方法中，合成單元的來源主要係將一套經過設計的錄音腳本，交由專業的錄音人士使用專業的錄音設備來錄製語音，之後再利用電腦系統根據上述錄音腳本中的標音資訊 (phonetic information)，以自動對錄音的音檔切音，進而擷取出合成系統所需的語音單元。
20

然而，由於習知電腦系統的切音位置並不保證完全準確，且數量龐大的錄音腳本需耗費相當長的時間來進行錄製作業，故即使是專業人士也可能發生音節沒唸好、漏唸(deletion)、多唸(insertion)、唸錯字、唸錯破音字、或因為唸太快所造成之連音效應(co-articulation)致使音節聽不清楚等情形。由於切音位置的準確性、以及合成單元的品質優劣及正確與否，將會直接影響語音合成的輸出品質，因此，如何提升電腦切音位置的可靠度，且正確地將一些沒錄製好的錄音語料篩選出來，再交由專業錄音人士重新錄製，即為語音合成系統中一極為重要的課題。

為檢查所錄製之合成單元是否正確，傳統係由人工操作工具軟體來逐一檢查每個合成單元。然而，由於現今合成系統大多採用數量龐大的合成單元，相對所需錄製的語料資料量也十分龐大，因此若使用人工方式來檢查所有的合成單元，不僅耗時費力，更容易受到個人主觀因素的影響而對同一組錄音語料的良窳判斷出迥異的結果，其無法遵循一套一致性的客觀標準來作業。因此，前述之習知技術實有予以改進之必要。

習知係提出相當多關於語音驗證之技術，例如揭露於美國專利公告第 6292778 號、美國專利公告第 6125345 號、美國專利公告第 5675706 號之專利文獻。其中，美國專利公告第 6292778 號專利係結合語音辨識器(speech recognizer)以及與文句無關語句驗證(task-independent utterance verifier)技術，來提升單詞(word)、片語(phrase)、

及句子(sentence)的辨識率，此驗證技術係利用驗證模型(model)及反向模型(anti-model)分別量測語音辨識結果的近似值(likelihood)，再取其對數概度比(log ratio)為分數，最後整合單詞、片語、及句子的分數與一事先定義好的門檻值比較，以決定欲拒絕或接受語音辨識結果；美國專利公告第 6125345 號專利則利用語音辨識器(speech recognizer)產生一個以上的可靠度量測，再將語音辨識結果傳給語音驗證器(recognition verifier)以產生一個以上之可靠度量測，最後透過整合器(integrator)利用多層感知器(multi-layer preceptron, MLP)來整合上述可靠度量測，以決定欲拒絕或接受語音辨識結果；而美國專利公告第 5675706 號專利則係利用單字(subword)及字串(string)兩個驗證階段來驗證一未知文字內容的語音段是否為關鍵字(keyword)，單字驗證階段檢驗隱藏式馬可夫模型(hidden Markov model, HMM)的辨識結果，字串驗證階段則整合單字驗證階段的結果，來決定拒絕或接受整個關鍵字。

然而，上述之專利文獻皆設計來解決語音辨識所面臨的問題，亦即用以辨認一個「未知」文字內容的語音段，迥異於大語料庫之語音合成技術所欲解決「已知」文字內容應用的課題；此外，語音辨識主要係欲解決集外詞(out of vocabulary, OOV)的問題，但大語料庫之語音合成技術則欲確認各已知語音單元之錄音及切音結果是否正確；又，語音辨識需要辨識的標的可能是一個字詞、一個片語、或是一句話，亦與大語料庫之語音合成技術將焦點放在一個

基本合成單位(例如一個音節)是否達到合成系統可採用之標準的應用不同。由此可知，習知所提出之語音辨識技術不論在條件或應用面上皆無法改善習知大語料庫語音合成系統的問題，並非十分理想。

5

【發明內容】

本發明之主要目的係在提供一種連續語音自動切音及驗證之方法及系統，其係同時整合切音與語音驗證之可靠度分析流程，俾能取得可靠的語音切點位置、同時篩選
10 出有問題的錄音語料，進而大幅提昇大語料庫合成單元收錄之方便性、正確性、以及品質。

本發明之另一目的係在提供一種連續語音自動切音及驗證之方法及系統，其係藉由電腦自動切割音檔來自動收集成成單元，俾取代由人工逐一檢查合成單元所造成的
15 疏忽及不一致性。

依據本發明之一特色，於所提出之連續語音自動切音及驗證之方法中，首先係擷取一錄音語料檔(recorded speech corpus)，此錄音語料檔係參照一已知文字腳本(known text script)所錄製而成，且已知文字腳本中係定義
20 有N個語音單元之語音資訊；接著將根據已知文字腳本中、各語音單元之語音資訊，而自錄音語料檔中對應切割出N個待測語音單元(test speech unit segment)；之後，將驗證待測語音單元之切點(cutting point)位置所對應之切音可靠度，藉以判斷上述待測語音單元所對應之切點位置

是否正確；且再驗證待測語音單元之語音可靠度，藉以判斷上述待測語音單元是否為根據已知文字腳本錄製而成；最後則結合待測語音單元所對應的切音可靠度及語音可靠度用以與預設門檻值比較，當可靠度大於門檻值時，表示
5 待測語音單元係被正確地錄製，故接受此待測語音單元，反之，則拒絕此待測語音單元。

依據本發明之另一特色，所提出之連續語音自動切音及驗證之系統主要包括有一資料庫、一單元切割模組、一切音驗證模組、一語音驗證模組、以及一單元篩選模組。
10 其中，資料庫中儲存有一已知文字腳本、以及參照已知文字腳本所錄製而成的錄音語料檔，且已知文字腳本中係已定義有N個語音單元之語音資訊；單元切割模組係根據已知文字腳本中各語音單元之語音資訊，自錄音語料檔中對應切割出N個待測語音單元；接下來將交由切音驗證模組
15 來驗證上述待測語音單元之切點位置所對應的可靠度，藉以判斷出待測語音單元所對應的切點位置是否正確；且亦需由語音驗證模組來驗證上述待測語音單元之語音可靠度，藉以判斷待測語音單元是否確實根據已知文字腳本錄製而成；待單元篩選模組接收到待測語音單元所對應之切
20 音可靠度及語音可靠度後，將結合前述可靠度資訊以與一預設門檻值相比，當可靠度大於門檻值時，表示此待測語音單元係被正確地錄製，故接受待測語音單元，反之，則拒絕待測語音單元。

【實施方式】

本發明所提出之連續語音自動切音及驗證之方法及系統係可應用在基於語料庫之文字轉語音 (corpus-based text-to-speech) 系統的合成語音單元建置，亦可應用在任何需要從大語料庫中切取語音單元的應用中。有關本發明之較佳實施例，請先參閱圖1之功能方塊圖，其顯示本實施例係使用於一電腦系統1以對參照已知文字腳本K(known text script)所錄製的錄音語料檔R(recorded speech corpus)來進行切音驗證及語音驗證。本實施例之電腦系統1包括有一資料庫11、一單元切割器12(speech unit segmentor)、一切音驗證器13(segmental verifier)、一語音驗證器14(phonetic verifier)、以及一單元篩選器15(speech unit inspector)，用以驗證自錄音語料檔R中所切割出的待測語音單元121之切音可靠度CMS及語音可靠度CMV。其中，資料庫11係用以儲存已知語音腳本K(又稱錄音腳本、或文字腳本)及錄音語料檔R，已知語音腳本K係定義有N個語音單元(speech unit segment)之語音資訊，錄音語料檔R較佳為委請專業錄音人士根據上述已知語音腳本K所錄製而成的真人語音。

由於下述之實施例係以中文(Mandarin)語音合成語料庫之應用為例，而中文係屬於單一音節(monosyllable)語言，亦即每一個中文字就代表一個音節(syllable)，因此在進行語音合成時所需使用的基本合成單元(即語音單元)較佳係為中文「音節」以便於舉例說明，由語音學上的觀點

來看，一個中文音節通常係由聲母(initial)及韻母(final)所組成，其中，聲母包括有鼻音(nasal)、流音(liquid)、塞音(stop)、摩擦音(fricative)、及破擦音(affricate)等音素(phonemic)種類。當然語音單元亦可以是單音(phone)、雙音(di-phone)、字根(root)、字首(prefix)、或詞(word)...等可供進行語音合成之語音單元，並不侷限於本實施例之應用範疇。

接下來請一併參閱圖2，其係詳述本實施例電腦系統1的驗證流程。首先，電腦系統1係接收由專業錄音人士根據已知文字腳本K所錄製而成的錄音語料檔R並將其儲存於資料庫11中(步驟S201)；接著，由單元切割器12根據預存於資料庫11中之已知文字腳本K的語音資料，以自錄音語料檔R的語音段中切割出N個待測語音單元121(text speech unit segment)(步驟S202)，藉以取得各個待測語音單元121的切點位置(cutting point)資訊。

為確保本實施例之單元切割器12所切割出的音節保有連續語音的詞語特性，因此請參閱圖3，其係詳述圖2之步驟S202中用以切割出待測語音單元121的細部流程，單元切割器12首先使用隱藏式馬可夫模型(hidden Markov model, HMM)對錄音語料檔R進行切音校準(alignment)(步驟S301)，以粗略地切割出N個待測語音單元121，且每一待測語音單元121皆對應有一初始切點位置，其中，隱藏式馬可夫模型使用之分析視窗的音框長度為20ms、音框位移為10ms，而切割出的每一特徵向量係具有26個維度的特徵

值，分別為12維的梅氏倒頻譜係數(Mel-cepstral coefficient)、12維的差分(delta)倒頻譜係數、與對數能量的1次及2次差分(delta-log-energy及delta-delta-log-energy)。此外，本實施例係使用不特定語者(speaker independent)之隱藏式馬可夫模型作為聲母模型(initial model)來訓練特定語者(speaker dependent)之隱藏式馬可夫模型。

接著，第二步將根據由隱藏式馬可夫模型所切割出的各個初始切點位置進行細部微調(fine adjustment)(步驟S302)，進而取得更精確的切點微調值，其包括根據不同的單元型態、語音參數之資料、還有搜尋單元等特徵參數來對初始切點位置進行微調，上述特徵參數例如為初始切點位置的相鄰(neighboring)切點位置、待測語音單元121之過零率(zero crossing rate, ZCR)、及待測語音之能量值(energy)，例如可使用音框長度為5ms、音框位移為1ms的視窗來計算出過零率及能量值以取得切點微調值，其中，能量值係為待測語音單元121之帶通訊號(band pass signal)及高通訊號(high pass signal)之能量值，其係擷取自一特定語者頻帶。

最後，由於本實施例所進行之語音切割係可應用於文字轉語音(text-to-speech, TTS)系統中，其重點在於提升切割音段的語音連續性，亦即不只需要正確的切割單元、更需要讓切割出的音節音調能夠接近一般常態，故第三步將整合上述待測語音單元121所對應的初始切點位置及切點

微調值，以決定出各待測語音的切點位置(步驟S303)。其中，整合方法例如可根據不同的類別而有不同的優先順序(priority)，也就是在某一個特定類別中、就尊重某一個專家意見；或可根據多個專家意見取平均值(average)；亦可
5 對各切點位置及切點微調值賦予權重平均值(weighted average)；當然，不同聲母的過零率統計值、或者各個不同聲母、韻母、及音節種類的週期統計值和能量統計值亦可作為整合的依據。

再請參考圖1之方塊圖及圖2之流程圖，其顯示在取得
10 待測語音單元121的切點位置資訊之後，電腦系統1將把待測語音單元121分別傳送至切音驗證器13和語音驗證器14來執行驗證流程，以確認錄音語料檔R中的音節和已知文字腳本K中的音節間的語音一致性(phonetic consistence)。切音驗證器13將驗證待測語音單元121的切音可靠度
15 CMS(步驟S203)，其係根據各種語音統計參數來判斷待測語音單元121所對應的切點位置是否正確，藉以決定出待測語音單元121的邊界(boundary)；而語音驗證器14則將驗證待測語音單元121的語音可靠度(步驟S204)，以判斷待測語音單元121是否確實根據已知文字腳本K錄製而成。需注意
20 的是，上述步驟S203及步驟S204較佳係為同時執行，當然亦可互調兩者之執行順序。

其中，由切音驗證器13計算出之切音可靠度CMS為：

$$CMS = \max\left(1 - h(D) - \sum_{s,f} g(c(s), f(s)), 0\right),$$

當中， $h(D) = K\left(\sum_i w_i |d_i - \bar{d}|\right)$ (或 $h(D) = K\left(\sum_i w_i (d_i - \bar{d})^2\right)$)， D 為待測

語音單元121之切點的多專家決策向量值， d_i 為切點位置，

$\bar{d} = p(D)$ 為切點之最後決策位置，其係可根據優先順序、平

均值、或權重值所決定出來， $K(x)$ 為一用以將非負值變數

5 對應(mapping)為介於0和1之間之數值的單調遞增函數
(monotonically increasing function)， $h(D)$ 係用以驗證在多

專家系統中、多個決策點(decision point)間的不一致性，

倘若不一致性越高，就表示可靠度越低， $g(c(s), f(s))$ 則為

一數值介於0和1之間的成本函數， s 為一音段， $c(s)$ 為音段 s

10 之類別， $f(s)$ 為音段 s 之聲學特徵，其可能是過零率、長度、

能量值、及週期...等。因此，本實施例之切音可靠度CMS

係介於0(可靠度最低)和1(可靠度最高)之間，而在計算過程

中，將會對不可靠的結果扣分，最低扣到0分為止，其中不

可靠的結果可能是切點位置係為發散不一致的情形...等。

15 舉例來說，若預期所切割出的待測語音單元121是一段摩擦

音，其對應之過零率應該略微偏高，但實際上卻切割出一

段沈默語音(silence)，則實際所測出的過零率將不若原先

預期的高，使得計算出的成本值變大，導致可靠度降低。

而本實施例之語音驗證器14則是根據語音辨識領域

20 中所採用的區辨詞語驗證(discriminative utterance

verification)來進行自動音節驗證程序，其針對每一音節類

別皆分別設計有一組音節模型(syllable model)和反向模型

(anti-model)以組成語音驗證模型(verification model)。音

節模型係訓練為用以辨識待測語音單元121和目標語音單

元類別相符的機率；反之，反向模型則訓練為用以辨識待測語音單元121不符合目標語音單元類別的機率，由於訓練音節模型與反向模型的方法係為此項技術領域之人士所熟知，故不在此贅述。據此，語音驗證器14將可計算出語音可靠度CMV為：

$$CMV = \min\{LLR_I, LLR_F, 0\},$$

當中， $\begin{cases} LLR_I = \log P(X_I | H_0) - \log P(X_I | H_1) \\ LLR_F = \log P(X_F | H_0) - \log P(X_F | H_1) \end{cases}$ ， X_I 為待測語音單元

121之聲母， X_F 為待測語音單元121之韻母， H_0 為待測語音單元121係被正確錄製之虛無假設(即對應於音節模型)， H_1 為待測語音單元121未被正確錄製之對立假設(即對應於反向模型)，LLR為一對數近似值比(log likelihood ratio)。

需注意的是，本實施例之切音驗證器13除可驗證切點位置之外，亦具有部分驗證語音可靠度的功能，因為倘若錄音語料檔R中所錄製的語音不正確時，所對應的特徵參數就會有誤，因此切音驗證器13可部分檢查出此錄音語料是否正確；此外，本實施例之語音驗證器14亦可在驗證語音內容之餘，具有驗證切點位置的功能，因為假如切點位置錯誤，則在進行語音驗證時將會受到切割錯誤的待測語音單元121影響、而產生較大的誤差，導致降低語音可靠度CMV。

最後，將由單元篩選器15結合切音可靠度CMS及語音可靠度CMV用以與一預設之門檻值比對(步驟S205)，藉以決定接受或拒絕待測語音單元121(步驟S206)。其中，單元

篩選器15係可採用前期決策(early decision)或後期決策(late decision)的作法來比對待測語音單元121的切音可靠度CMS及語音可靠度CMV。而前期決策方式係可區分為兩種作法，第一種是切音可靠度CMS及語音可靠度CMV皆分別與門檻值比較後，判斷出兩者皆為可靠時，才接受待測語音單元121以收錄至大語料庫中；第二種則是切音可靠度CMS或語音可靠度CMV其中之一為可靠的，就接受待測語音單元121。至於後期決策方式，則是在正規化切音可靠度CMS及語音可靠度CMV後，賦予各可靠度不同的權重值以計算出一單一可靠度用以和門檻值比較來作決策。

由以上說明可知，本發明可透過電腦自動切割音檔、及整合切音與語音驗證之可靠度分析的流程，除了可取得可靠的切割位置、且檢驗出有問題的錄音語料外，更可避免因人工檢查所造成的疏忽及不一致性，因此大大提升了大語料庫語音單元收錄的方便性、正確性、及品質。

上述實施例僅係為了方便說明而舉例而已，本發明所主張之權利範圍自應以申請專利範圍所述為準，而非僅限於上述實施例。

20 【圖式簡單說明】

圖1係本發明一較佳實施例之功能方塊圖。

圖2係本發明一較佳實施例之流程圖。

圖3係本發明一較佳實施例之切音步驟之流程圖。

【圖號說明】

電腦系統 1	資料庫 11
單元切割器 12	待測語音單元 121
切音驗證器 13	語音驗證器 14
5 單元篩選器 15	切音可靠度 CMS
語音可靠度 CMV	已知文字腳本 K
錄音語料檔 R	

拾、申請專利範圍：

1. 一種連續語音自動切音及驗證之方法，包括下列步驟：

5 一擷取錄音語料檔步驟，係擷取一錄音語料檔，該錄音語料檔之內容係參照一已知文字腳本所錄製而成，該已知文字腳本中係定義有N個語音單元之語音資訊；

一切音步驟，係根據該已知文字腳本中該等語音單元之語音資訊，自該錄音語料檔中對應切割出N個待測語音單元；

10 一切音驗證步驟，係驗證該等待測語音單元之切點位置所對應之切音可靠度，以判斷該等待測語音單元所對應之切點位置是否正確；

一語音驗證步驟，係驗證該等待測語音單元之語音可靠度，以判斷該等待測語音單元是否係為根據該已知文字腳本錄製而成；以及

15 一決策步驟，係結合該待測語音單元所對應之切音可靠度及語音可靠度用以與一預設之門檻值比對，當該可靠度大於該門檻值時，接受該待測語音單元，反之，則拒絕該語音單元。

20 2. 如申請專利範圍第2項所述之方法，其中，該切音步驟係包括下列步驟：

根據該已知文字腳本中該等語音單元之語音資訊，使用隱藏式馬可夫模型以自該錄音語料檔中對應切割出N個

待測語音單元，其中每一待測語音單元係對應定義有一初始切點位置；

- 根據各待測語音單元所對應之至少一特徵參數，以對該待測語音單元之初始切點位置進行微調，並計算出該待測語音單元對應之至少一切點微調值；以及

整合該待測語音單元之初始切點位置及切點微調值以取得該待測語音單元之切點位置。

3. 如申請專利範圍第2項所述之方法，其中，該待測語音單元所對應之特徵參數係為該初始切點位置之相鄰切點位置。

4. 如申請專利範圍第2項所述之方法，其中，該待測語音單元之特徵參數係為該待測語音單元之過零率。

5. 如申請專利範圍第2項所述之方法，其中，該待測語音單元之特徵參數係為該待測語音單元之能量值。

6. 如申請專利範圍第5項所述之方法，其中，該能量值係為該待測語音單元之帶通訊號及高通訊號之能量值，其係擷取自一特定語者頻帶。

7. 如申請專利範圍第2項所述之方法，其中，每一切點微調值係定義有一權重值，該待測語音單元之切點位置係為該初始切點位置及切點微調值之權重平均值。

8. 如申請專利範圍第1項所述之方法，其中，於該切音驗證步驟中，該待測語音單元所對應之切音可靠度係為：

$$CMS = \max \left(1 - h(D) - \sum_{s,f} g(c(s), f(s)), 0 \right),$$

當中， $h(D) = K \left(\sum_i w_i |d_i - \bar{d}| \right)$ ， D 為該待測語音單元之切點之多
專家決策向量值， d_i 為切點位置， $\bar{d} = p(D)$ 為該切點之最後
決策位置， $K(x)$ 為一用以將非負值變數對應為介於0和1之
間之數值的單調遞增函數， $g(c(s), f(s))$ 為一數值介於0和1
5 之間的成本函數， s 為一音段， $c(s)$ 為音段 s 之類別， $f(s)$ 為
音段 s 之聲學特徵。

9. 如申請專利範圍第1項所述之方法，其中，於該語
音驗證步驟中，該待測語音單元所對應之語音可靠度係為：

$$CMV = \min\{LLR_I, LLR_F, 0\},$$

10 當中， $\begin{cases} LLR_I = \log P(X_I | H_0) - \log P(X_I | H_1) \\ LLR_F = \log P(X_F | H_0) - \log P(X_F | H_1) \end{cases}$ ， X_I 為該待測語音單

元之聲母， X_F 為該待測語音單元之韻母， H_0 為該待測語音
單元係被正確錄製之虛無假設， H_1 為該待測語音單元未被
正確錄製之對立假設， LLR 為一對數近似值比。

10. 一種連續語音自動切音及驗證之系統，包括：

15 一資料庫，係儲存有一已知文字腳本、以及一參照該
已知文字腳本所錄製而成之錄音語料檔，該已知文字腳本
中係定義有 N 個語音單元之語音資訊(N 為正整數)；

一單元切割模組，根據該已知文字腳本中該等語音單
元之語音資訊，自該錄音語料檔中對應切割出 N 個待測語
20 音單元；

一切音驗證模組，係驗證該等待測語音單元之切點位
置所對應之切音可靠度，以判斷該等待測語音單元所對應
之切點位置是否正確；

一語音驗證模組，係驗證該等待測語音單元之語音可靠度，以判斷該等待測語音單元是否係為根據該已知文字腳本錄製而成；以及

- 5 一單元篩選模組，係結合該待測語音單元所對應之切音可靠度及語音可靠度用以與一預設之門檻值比對，當該可靠度大於該門檻值時，接受該待測語音單元，反之，則拒絕該語音單元。

11. 如申請專利範圍第10項所述之系統，其中，該切音驗證模組係根據下列步驟以驗證該等待測語音單元之切點位置所對應之切音可靠度：

根據該已知文字腳本中該等語音單元之語音資訊，使用隱藏式馬可夫模型以自該錄音語料檔中對應切割出N個待測語音單元，其中每一待測語音單元係對應定義有一初始切點位置；

- 15 根據各待測語音單元所對應之至少一特徵參數，以對該待測語音單元之初始切點位置進行微調，並計算出該待測語音單元對應之至少一切點微調值；以及

整合該待測語音單元之初始切點位置及切點微調值以取得該待測語音單元之切點位置。

- 20 12. 如申請專利範圍第11項所述之系統，其中，該待測語音單元所對應之特徵參數係為該初始切點位置之相鄰切點位置。

13. 如申請專利範圍第11項所述之系統，其中，該待測語音單元之特徵參數係為該待測語音單元之過零率。

14. 如申請專利範圍第11項所述之系統，其中，該待測語音單元之特徵參數係為該待測語音單元之能量值。

15. 如申請專利範圍第14項所述之系統，其中，該能量值係為該待測語音單元之帶通訊號及高通訊號之能量值，
5 其係擷取自一特定語者頻帶。

16. 如申請專利範圍第11項所述之系統，其中，每一切點微調值係定義有一權重值，該待測語音單元之切點位置係為該初始切點位置及切點微調值之權重平均值。

17. 如申請專利範圍第10項所述之系統，其中，該待測
10 語音單元所對應之切音可靠度係為：

$$CMS = \max\left(1 - h(D) - \sum_{s,f} g(c(s), f(s)), 0\right),$$

當中， $h(D) = K\left(\sum_i w_i |d_i - \bar{d}|\right)$ ， D 為該待測語音單元之切點之多
專家決策向量值， d_i 為切點位置， $\bar{d} = p(D)$ 為該切點之最後
決策位置， $K(x)$ 為一用以將非負值變數對應為介於0和1之
15 間之數值的單調遞增函數， $g(c(s), f(s))$ 為一數值介於0和1
之間的成本函數， s 為一音段， $c(s)$ 為音段 s 之類別， $f(s)$ 為
音段 s 之聲學特徵。

18. 如申請專利範圍第10項所述之系統，其中，該待測語音單元所對應之語音可靠度係為：

20 $CMV = \min\{LLR_I, LLR_F, 0\},$

當中， $\begin{cases} LLR_I = \log P(X_I | H_0) - \log P(X_I | H_1) \\ LLR_F = \log P(X_F | H_0) - \log P(X_F | H_1) \end{cases}$ ， X_I 為該待測語音單

元之聲母， X_F 為該待測語音單元之韻母， H_0 為該待測語音

單元係被正確錄製之虛無假設， H_1 為該待測語音單元未被
正確錄製之對立假設，LLR為一對數近似值比。

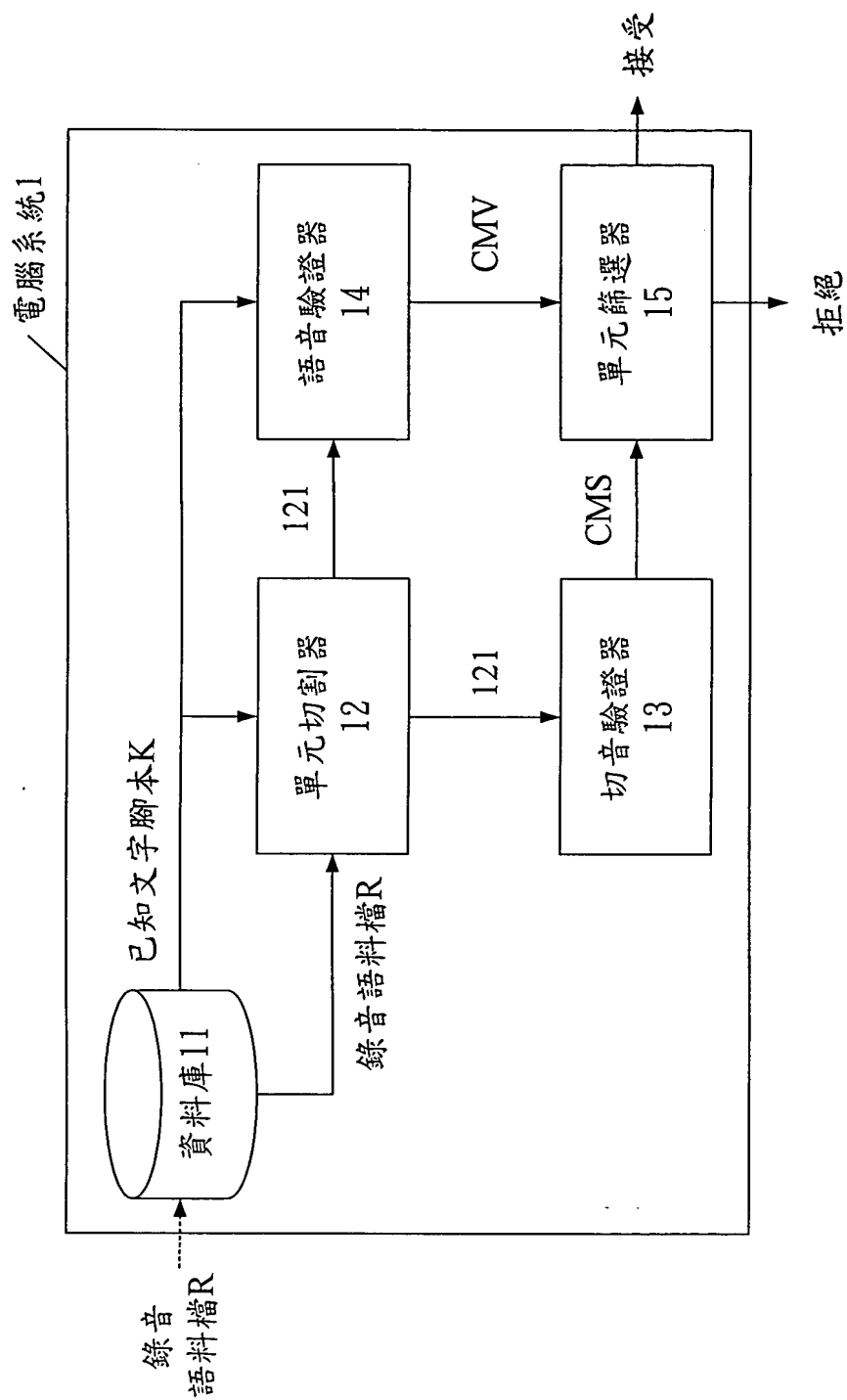


圖1

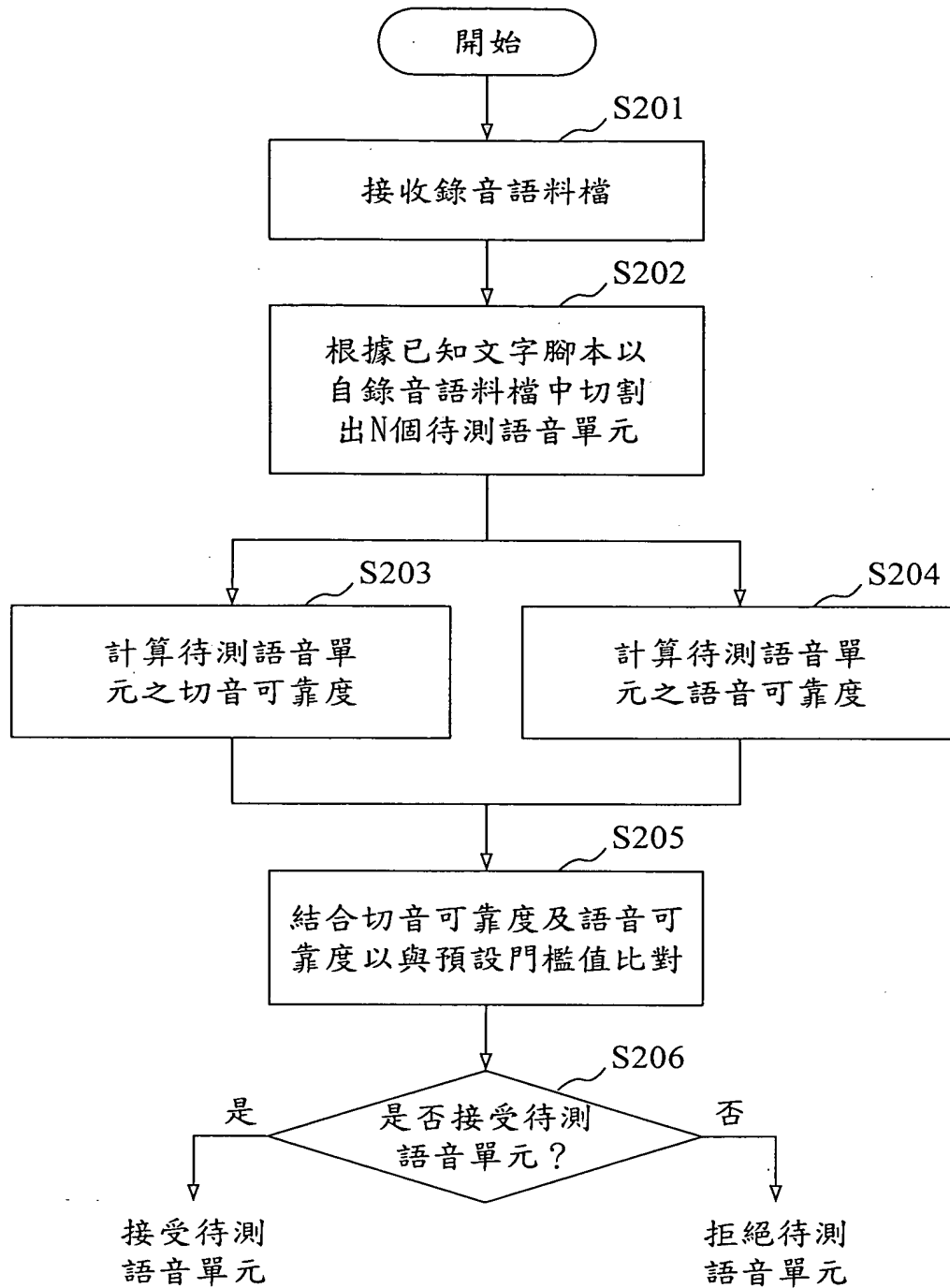


圖2

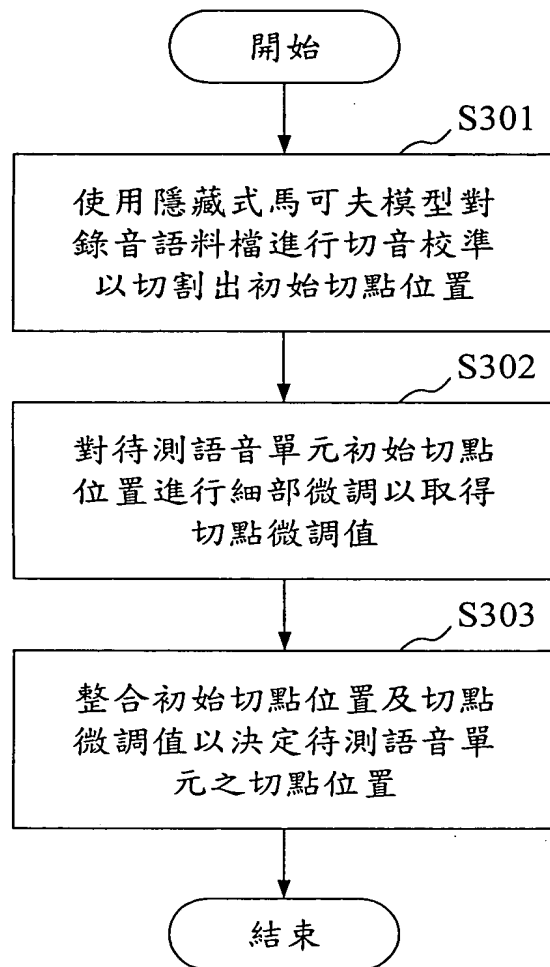


圖3